Oja Medians and Centers of Gravity

Dan Chen Olivier Devillers John Iacono Stefan Langerman Pat Morin

Abstract

The relationships between the Oja median and centers of gravity are considered.

1 Introduction

Given a set S of n points in $\mathbb{R}^d$, the *Oja depth* [7] of a point $x \in \mathbb{R}^d$ is

$$\mathrm{d}(x,S) = \sum_{y_1,\ldots,y_d \in \binom{S}{d}} \mathrm{v}(x, y_1, \ldots, y_d) \ ,$$

where $\mathrm{v}(p_1, \ldots, p_{d+1})$ denotes the volume of the simplex whose vertices are $p_1 \ldots p_{d+1}$.[1] A point in $\mathbb{R}^d$ with the minimum Oja depth is called an *Oja center*.

1.1 New Results

In this paper we consider relationships between centers of gravity of certain sets and Oja depth. The *center of gravity* of a finite point set $S \subset \mathbb{R}^d$ is the average of those points,

$$\mathrm{c}(S) = |S|^{-1} \sum_{x \in S} x \ .$$

If $P \subset \mathbb{R}^d$ is a bounded object of non-zero volume, the center of gravity of P is

$$\mathrm{c}(P) = \frac{\int_{x \in P} x \, dx}{\mathrm{v}(P)} \ .$$

In this paper, we prove the following results about the Oja depth of an n point set S, whose convex hull A has unit volume and that has an Oja center x:

$$\mathrm{d}(\mathrm{c}(A), S) \le \binom{n}{d} / (d+1) \ , \tag{1}$$

$$\mathrm{d}(\mathrm{c}(S), S) \le (d+1)\, \mathrm{d}(x, S) \ . \tag{2}$$

1.2 Related Results

Our first result, (1), is a form of *Centerpoint Theorem* that upper-bounds the Oja depth of $\mathrm{c}(A)$, and hence also the Oja depth of x, in terms of the volume of the convex hull of S. Previously, centerpoint theorems were known for other depth functions such as Tukey depth [5, 8, 10] and simplicial depth [2, 3, 4]. To the best of our knowledge, this is the first such result for Oja depth.

Our next result, (2), can be viewed in two ways. The first is as a linear-time constant factor approximation for finding an Oja median.

In 1-d, Oja depth is minimized by the median, which can be found in $O(n)$ time. However, in 2-d, the best known algorithm for minimizing Oja depth exactly takes $O(n \log^3 n)$ time [1]. Approximation algorithms for minimizing Oja depth, based on uniform grids and sampling from $\binom{S}{d}$, are given by Ronkainen, Oja, and Orponen [9]. However, in pathological cases, their approximation algorithm is not guaranteed (or even likely) to find a point that closely approximates the Oja median, either in terms of distance or in terms of its Oja depth.[2]

Another view of (2) is that it gives insight into the Oja depth function and the Oja median. In some sense, it tells us that the Oja median is not terribly different from the center of gravity of S, since the center of gravity of S minimizes, to within a constant factor, the Oja depth function.

2 Oja Center and Gravity Center of A

In this section, we relate the Oja depth of the center of gravity of the convex hull of S to the volume of the convex hull of S. Throughout this section, A denotes the convex hull of S and we assume, without loss of generality, that $\mathrm{v}(A) = 1$.

Our upper-bound is based on the following identity: For any disjoint sets $X, Y \subseteq \mathbb{R}^d$,

$$\mathrm{c}(X \cup Y) = \frac{\mathrm{v}(X)\,\mathrm{c}(Y) + \mathrm{v}(Y)\,\mathrm{c}(Y)}{\mathrm{v}(X \cup Y)}.$$

First let us introduce a notion from convex geometry. Let A be a convex body in $\mathbb{R}^d$, where $d \ge 2$. Suppose A lies between parallel hyperplanes $x_1 = a$ and $x_1 = b$, where $a < b$. For each x with $a \le x \le b$, let A_x be the intersection of A with the hyperplane $x_1 = x$, and define r_x by the equation

$$\omega_{d-1} r_x^{d-1} = \mathrm{v}_{d-1}(A_x),$$

where $\mathrm{v}_{d-1}(X)$ denotes the $(d-1)$-dimensional volume of X and ω_{d-1} is the $(d-1)$-dimensional volume of the

[1] In Oja's original definition, the sum is normalized by dividing by $\binom{|S|}{d}$. We omit this here since it changes none of our results and clutters our formulas.

[2] This follows from the fact that the value of the Oja depth function and the location of the Oja median can be arbitrarily different for two sets S_1 and S_2 that differ in only d points [6].

unit $(d-1)$-ball. In this way, r_x is the radius of a $(d-1)$-ball whose v_{d-1}-volume is the same as that of A_x. For each $a \leq x \leq b$, let C_x be defined by the equation

$$C_x = \{(x, x_2, \ldots, x_d) : x_2^2 + \cdots + x_d^2 \leq r_x^2\}.$$

Then the set

$$C = \cup(C_x : a \leq x \leq b)$$

is called the *Schwarz rotation-symmetral* of A in the x_1-axis. For example, in $\mathbb{R}^3$, C is a stack of disks perpendicular to, and centered on, the x-axis. Each disk has the same area as the corresponding slice of A.

Theorem 1 (Webster [11]) *Let A be a convex body in $\mathbb{R}^d$ $(d \geq 2)$ whose Schwarz rotation-symmetral in the x_1-axis is C. Then C is a convex body having the same volume as A.*

Lemma 2 *Let g be the center of gravity of a convex d-polytope P. Then any d-simplex T whose vertices are g plus d points inside P has volume at most $\mathrm{v}(P)/(d+1)$.*

Proof. Let $p_1, \ldots, p_d$ be any d points in P, and let h be the hyperplane that contains them. If g is in h, then $\mathrm{v}(T) = 0$. If not, rotate P to make h perpendicular to the x_1-axis with g above h. If there is any volume of P below h, we can cut that part off from P to obtain a new polytope P'. The volume of P' will be less than 1, and its gravity center g' will be above g. In this way, if $(P, p_1, \ldots, p_d)$ is a counterexample to the lemma, then so is $(P', p_1, \ldots, p_d)$. The face of P' in h is a convex hull B containing the d points. Let q be a point above h such that the pyramid D with B as base and q as apex has the same volume as P'.

Let C be the Schwarz rotation-symmetral of P' in the x_1-axis, and R be that of D (see Figure 1). Note that R is a conic pyramid. By Theorem 1, C is convex and $\mathrm{v}(C) = \mathrm{v}(P') = \mathrm{v}(R)$. Let c be the intersection of the surfaces of C and R above B. Note that the surface of R is bounded by a collection of lines that pass through q. Each of these lines intersects C in at most 2 points. One of these points has the same x_1-coordinate as B and the other points lie on the boundary of a $(d-1)$-ball c.

Let x_c be the x_1-coordinate of c. Since C is convex. Then the surface of C below $x_1 = x_c$ is outside the surface of R. By the definition of Schwarz rotation-symmetral, the volume of C that is outside of R is below x_c, and the volume of R outside of C is above x_c. Therefore, the gravity center of R is above that of C because of central identity.

The gravity centers of P' and C have the same height because in the Schwarz rotation-symmetral C_x has the same x_1 value as A_x. So do the gravity centers of D and R. Let g_d be gravity center of D. Since D is a pyramid,

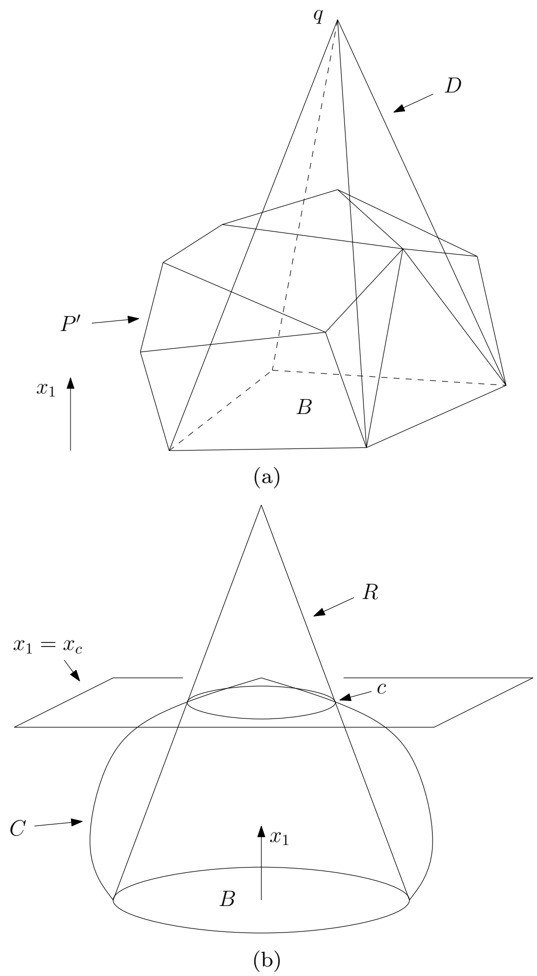

Figure 1: The Schwarz rotation-symmetral of P' and D

the convex hull of the d points is contained in B, g is below g', and g' is below g_d, so

$$\mathrm{v}(T) \leq \mathrm{v}(P')/(d+1) \leq \mathrm{v}(P)/(d+1).$$

To see this, consider that $\mathrm{v}(T) = \mathrm{v}(g, p_1, \ldots, p_d) \leq \mathrm{v}(g_d, p_1, \ldots, p_d) = \mathrm{v}(q, p_1, \ldots, p_d)/(d+1)$. $\square$

Theorem 3 *Let S be a set of points in $\mathbb{R}^d$ whose convex hull, A, has unit volume. Then $\mathrm{d}(\mathrm{c}(A), S) \leq \binom{n}{d}/(d+1)$.*

Proof. According to Lemma 2,

$$\begin{aligned}\mathrm{d}(\mathrm{c}(A), S) &= \sum_{y_1,\ldots,y_d \in \binom{S}{d}} \mathrm{v}(\mathrm{c}(A), y_1, \ldots, y_d) \\ &\leq \binom{n}{d}/(d+1).\end{aligned}$$

$\square$

3 Oja Center and Gravity Center of S

In this section, we show that the center of gravity of S provides a constant-factor approximation to the point of minimum Oja depth.

Theorem 4 *For any finite set $S \subset \mathbb{R}$, $\mathrm{d}(\mathrm{c}(S), S) \leq 2\,\mathrm{d}(x, S)$ for any $x \in \mathbb{R}$.*

Proof. Denote the elements of S by $p_1, \ldots, p_n$ in any order. Let the multiset S_i contain $p_1, \ldots, p_i$ as well as $n - i$ copies of x. Let $c_i = \mathrm{c}(S_i)$. We will show, by induction on i, that $\mathrm{d}(c_i, S_i) \leq 2\,\mathrm{d}(x, S_i)$ for all $i \in \{0, \ldots, n\}$. This is sufficient, since $S_n = S$.

For the base case S_0 consists of n copies of x, so $c_0 = x$ and $\mathrm{d}(c_0, S_0) = 0 = 2\,\mathrm{d}(x, S_0)$. Next, we assume that $\mathrm{d}(c_i, S_i) \leq 2\,\mathrm{d}(x, S_i)$ and prove that $\mathrm{d}(c_{i+1}, S_{i+1}) \leq 2\,\mathrm{d}(x, S_{i+1})$. Note that

$$\mathrm{d}(x, S_{i+1}) = \mathrm{d}(x, S_i) + |p_{i+1} - x| \ .$$

Furthermore,

$$c_{i+1} = c_i + (p_{i+1} - x)/n \ ,$$

so

$$\begin{aligned}
\mathrm{d}(c_{i+1}, S_{i+1}) = {} & \mathrm{d}(c_i, S_i) \\
& + \sum_{q \in S_i} (|c_{i+1} - q| - |c_i - q|) \\
& + (|c_{i+1} - p_{i+1}| - |c_{i+1} - x|) \\
\leq {} & \mathrm{d}(c_i, S_i) \quad + n|p_{i+1} - x|/n \\
& + (|c_{i+1} - p_{i+1}| - |c_{i+1} - x|) \\
\leq {} & \mathrm{d}(c_i, S_i) + 2|p_{i+1} - x| \\
\leq {} & 2\,\mathrm{d}(x, S_i) + 2|p_{i+1} - x| \\
= {} & 2\,\mathrm{d}(x, S_{i+1}) \ ,
\end{aligned}$$

as required. $\square$

We remark that the above proof uses little more than triangle inequality. In particular, the same proof shows that the center of gravity gives a 2-approximation for the Fermat-Weber center in any dimension.[3] Unfortunately, in higher dimensions, Oja depth does not enjoy this nice property.

Theorem 5 *For any finite set $S \subseteq \mathbb{R}^d$, $\mathrm{d}(\mathrm{c}(S), S) \leq (d+1)\,\mathrm{d}(x, S)$ for any $x \in \mathbb{R}^d$.*

Proof. In this proof, we will make use of the fact that, for any d-simplex T with vertex set V_T and a point $q \in \mathbb{R}^d$,

$$\mathrm{v}(T) \leq \sum_{p_1, \ldots, p_d \in \binom{V_T}{d}} \mathrm{v}(p_1, \ldots, p_d, q) \ , \tag{3}$$

since T is contained in the union of the simplices on the right hand side. Equality occurs if q is inside T.

[3]The Fermat-Weber center of a point set S in $\mathbb{R}^d$ is the point x that minimizes $\sum_{y \in S} \|x - y\|$.

Define S_i as in the proof of Theorem 4. Let $S' = S_{i+1} \setminus \{p_{i+1}\}$. The induction and base case are the same as in Theorem 4. First, we have

$$\begin{aligned}
\mathrm{d}(x, S_{i+1}) \quad = \quad & \mathrm{d}(x, S_i) \\
& + \sum_{Q \in \binom{S_i}{d-1}} \mathrm{v}(x, p_{i+1}, Q),
\end{aligned} \tag{4}$$

where Q is a set of $d-1$ points, and

$$\begin{aligned}
& \mathrm{d}(c_{i+1}, S_{i+1}) \\
& \quad = \mathrm{d}(c_i, S_i) \\
& \qquad + \sum_{P \in \binom{S_i}{d}} (\mathrm{v}(c_{i+1}, P) - \mathrm{v}(c_i, P)) && (5) \\
& \qquad + \sum_{Q \in \binom{S'}{d-1}} (\mathrm{v}(c_{i+1}, p_{i+1}, Q) - \mathrm{v}(c_{i+1}, x, Q)), && (6)
\end{aligned}$$

where P is a set of d points. We denote $y^\perp$ the projection of a point y on a line perpendicular to the $d-1$ dimensional simplex P.

$$\begin{aligned}
& |\,\mathrm{v}(c_{i+1}, P) - \mathrm{v}(c_i, P)| \\
& \quad = \frac{1}{d}\,\mathrm{v}_{d-1}(P)\ |\ ||c_i^\perp P^\perp|| - ||c_{i+1}^\perp P^\perp||\ | \\
& \quad \leq \frac{1}{d}\,\mathrm{v}_{d-1}(P)||c_i^\perp c_{i+1}^\perp|| \\
& \quad \leq \frac{1}{d}\,\mathrm{v}_{d-1}(P)||\frac{1}{n} x^\perp p_{i+1}^\perp||
\end{aligned}$$

Then if $x^\perp$ and $p_{i+1}^\perp$ are on the same side of the hyperplane supporting P, we have

$$\begin{aligned}
& \frac{1}{d}\,\mathrm{v}_{d-1}(P)||\frac{1}{n} x^\perp p_{i+1}^\perp|| \\
& \quad \leq \frac{1}{nd}\,\mathrm{v}_{d-1}(P)\ \left|\ ||x^\perp P^\perp|| - ||p_{i+1}^\perp P^\perp||\ \right| \\
& \quad \leq \frac{1}{n}|\,\mathrm{v}(p_{i+1}, P) - \mathrm{v}(x, P)| \\
& \quad \leq \frac{1}{n} \sum_{Q \in \binom{P}{d-1}} \mathrm{v}(x, p_{i+1}, Q)
\end{aligned}$$

Ohterwise if $x^\perp$ and $p_{i+1}^\perp$ are on different sides of the hyperplane supporting P, we have $||x^\perp p_{i+1}^\perp|| = ||x^\perp P^\perp|| + ||p_{i+1}^\perp P^\perp||$. In this case the two simplices Px and Pp_{i+1} are disjoints and the convex hull of Pxp_{i+1} is covered by the union of the simplices Qxp_{i+1} for $Q \in \binom{P}{d-1}$, thus

$$\begin{aligned}
& \frac{1}{d}\,\mathrm{v}_{d-1}(P)||\frac{1}{n} x^\perp p_{i+1}^\perp|| \\
& \quad \leq \frac{1}{nd}\,\mathrm{v}_{d-1}(P)(||x^\perp P^\perp|| + ||p_{i+1}^\perp P^\perp||) \\
& \quad \leq \frac{1}{n}\,\mathrm{v}(p_{i+1}, P) + \mathrm{v}(x, P) \\
& \quad \leq \frac{1}{n} \sum_{Q \in \binom{P}{d-1}} \mathrm{v}(x, p_{i+1}, Q)
\end{aligned}$$

Then we can have (5)≤(4) as the follows:

$$\begin{aligned}
&\sum_{P\in\binom{S_i}{d}} (\mathrm{v}(c_{i+1},P)-\mathrm{v}(c_i,P)) \\
&\le \sum_{P\in\binom{S_i}{d}} (|\,\mathrm{v}(c_{i+1},P)-\mathrm{v}(c_i,P)|) \\
&\le \frac{1}{n}\sum_{P\in\binom{S_i}{d}}\sum_{Q\in\binom{P}{d-1}} \mathrm{v}(x,p_{i+1},Q) \\
&\le \frac{n-(d-1)}{n}\sum_{Q\in\binom{S_i}{d-1}} \mathrm{v}(x,p_{i+1},Q)\ .
\end{aligned}$$

Next, we show that $(6)\le d\times(4)$. Applying (3),

$$\begin{aligned}
&\sum_{Q\in\binom{S'}{d-1}} (\mathrm{v}(c_{i+1},p_{i+1},Q)-\mathrm{v}(c_{i+1},x,Q)) \\
&\le \sum_{Q\in\binom{S'}{d-1}}\left(\mathrm{v}(x,p_{i+1},Q)+\sum_{R\in\binom{Q}{d-2}}\mathrm{v}(x,p_{i+1},c_{i+1},R)\right) \\
&\le \sum_{Q\in\binom{S_i}{d-1}} \mathrm{v}(x,p_{i+1},Q) \\
&\quad +(n-1-(d-2))\sum_{R\in\binom{S_i}{d-2}} \mathrm{v}(x,p_{i+1},c_{i+1},R)\ ,
\end{aligned}$$

where R is a set of $d-2$ points. By linearity of determinant we have

$$\begin{aligned}
\bar{\mathrm{v}}(x,p_{i+1},c_{i+1},R) &= \frac{1}{n}\sum_{y\in S_{i+1}}\bar{\mathrm{v}}(x,p_{i+1},y,R) \\
&= \frac{1}{n}\sum_{y\in S_i}\bar{\mathrm{v}}(x,p_{i+1},y,R)
\end{aligned}$$

Since the absolute value of the sum can be bounded by the sum of the absolute values, we get

$$\mathrm{v}(x,p_{i+1},c_{i+1},R)\le\frac{1}{n}\sum_{y\in S_i}\mathrm{v}(x,p_{i+1},y,R),$$

and thus

$$\sum_{R\in\binom{S_i}{d-2}}\mathrm{v}(x,p_{i+1},c_{i+1},R)=\frac{d-1}{n}\sum_{Q\in\binom{S_i}{d-1}}\mathrm{v}(x,p_{i+1},Q).$$

Thus we can get $(6)\le d\times(4)$.

Finally, we resubstitute to obtain

$$\begin{aligned}
&\mathrm{d}(c_{i+1},S_{i+1}) \\
&\quad\le \mathrm{d}(c_i,S_i)+(d+1)\sum_{Q\in\binom{S_i}{d-1}}\mathrm{v}(x,p_{i+1},Q) \\
&\quad\le (d+1)\,\mathrm{d}(x,S_i)+(d+1)\sum_{Q\in\binom{S_i}{d-1}}\mathrm{v}(x,p_{i+1},Q) \\
&\quad= (d+1)\,\mathrm{d}(x,S_{i+1})\ .
\end{aligned}$$

Then we have $\mathrm{d}(\mathrm{c}(S),S)\le(d+1)\,\mathrm{d}(x,S)$. □

We remark that Theorem 4 and 5 are essentially the best possible. To see this, take the multiset S that contains $n-d$ copies of the origin o, and each of the remaining d points has one different coordinate 1 and all other coordinates 0. In this case $\mathrm{d}(o,S)=1/d!$ and $\mathrm{d}(\mathrm{c}(S),S)=(d+1-O(d^2/n))\times 1/d!$.

4 Conclusion

We have given several results on Oja depth and centers of gravity. There are several directions for future work.

Theorem 3 has no matching lower bound. The best lower-bound we know is that placing $n/(d+1)$ points at each vertex of any d-simplex of unit volume yields to an Oja depth of $n^d/(d+1)^d$ for any point inside the simplex. For $d=2$, for example, Theorem 3 implies $\mathrm{d}(x,S)\le n^2/6-O(n)$ where as the best lower bound (above) has $\mathrm{d}(x,S)\ge n^2/9$. This construction leads us to our first conjecture:

Conjecture 1 *For any point set $S\subset\mathbb{R}^d$ whose convex hull has unit volume, there exists $x\in\mathbb{R}^d$, such that $\mathrm{d}(x,S)\le n^d/(d+1)^d$*

References

[1] G. Aloupis, S. Langerman, M. Soss, and G. Toussaint. Algorithms for bivariate medians and a Fermat-Toricelli problem for lines. *Computational Geometry: Theory and Applications*, 26(1):69–79, 2003.

[2] I. Bárány. A generalization of Carathéodory's Theorem. *Discrete Mathematics*, 40:141–150, 1982.

[3] E. Boros and Z. Füredi. The maximal number of covers by the triangles of a given vertex set in the plane. *Geometrica Dedicata*, 17:69–77, 1984.

[4] R. Liu. On a notion of data depth based upon random simplices. *The Annals of Statistics*, 18:405–414, 1990.

[5] J. Matoušek. *Lectures in Discrete Geometry*. Springer-Verlag, New York, NY, 2002.

[6] A. Niinimaa, H. Oja, and M. Tableman. The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statistics & Probability Letters*, 10:325–328, 1990.

[7] H. Oja. Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1(6):327–332, 1983.

[8] J. Pach and P. K. Agarwal. *Combinatorial Geometry*. John Wiley & Sons, New York, NY, 1995.

[9] T. Ronkainen, H. Oja, and P. Orponen. Computation of the multivariate Oja median. In R. Dutter and P. Filzmoser, editors, *International Conference on Robust Statistics (ICORS 2001)*, 2003.

[10] J. Tukey. Mathematics and the picturing of data. In *International Conference of Mathematicians*, 1971.

[11] R. Webster. *Convexity*. Oxford University Press, New York, USA, 1995.